Consciousness and Cognition 18 (2009) 414–427

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Speculations on the emergence of self-awareness in big-brained organisms: The roles of associative memory and learning, existential and religious questions, and the emergence of tautologies

Emmanuel Tannenbaum

Department of Chemistry, Ben-Gurion University of the Negev, Sderot Ben-Gurion, Be'er-Sheva, Israel

ARTICLE INFO

Article history:
Received 2 April 2008
Available online 14 July 2008

Keywords:
Self-awareness
Consciousness
Associative memory
Associative learning
Tautologies
Self-image

ABSTRACT

This paper argues that self-awareness emerges in organisms whose brains have a sufficiently integrated, complex ability for associative learning and memory. Continual sensory input of information related to the organism leads to the formation of a set of associations that may be termed an organismal "self-image". After providing the basic mechanistic basis for the emergence of an organismal self-image, this paper proceeds to go through a representative list of behaviors associated with self-awareness, and shows how associative memory and learning, combined with an organismal self-image, leads to the emergence of these various behaviors. This paper also discusses various tautologies that invariably emerge when discussing self-awareness. We continue with various speculations on manipulating self-awareness, and discuss how concepts from set and logic theory may provide a useful set of tools for understanding the emergence of higher cognitive functions in complex organisms.

1. Introduction

Self-awareness is one of the most mysterious phenomenon in the natural world. Inanimate matter, through the process of replicative selection, has managed to produce biochemical machines that can recognize themselves. The existence of self-awareness has provoked much philosophical debate over the centuries, and, within a theological context, is intimately connected with the existence of a metaphysical soul.

While it was once widely believed that only humans were truly self-aware, it was later discovered that the great apes could recognize their own reflections, and therefore had a much higher level of self-awareness than previously thought. It was later discovered that dolphins also exhibit this level of self-awareness. Recently, elephants have been added to the list of animals capable of recognizing their own reflections (Roth & Dicke, 2005; Reiss & Marino, 2001; Plotnik, de Waal, & Reiss, 2006).

A major difficulty with understanding self-awareness has been its theological association with the concept of a soul. Because the phenomenon was regarded as mysterious, there have been relatively few scientific attempts to attack the problem. It was essentially viewed as an epiphenomenon associated with a sufficiently developed intelligence (Harter, 1999; Levenson, 2001).

A mechanistic study of self-awareness is possible, if we abandon traditional philosophical attempts to define what it is (which is a scientifically meaningless question, in any event), and instead attempt to study its operational aspects. According to Turing, a mechanistic approach to self-awareness requires us to determine what behaviors we would associate with

E-mail address: emanuelt@bgu.ac.il

1053-8100/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2008.06.003

self-awareness, so that self-awareness, in a precise sense, becomes defined by these behaviors (Korukonda, 2003). The next step is to then determine what is the mechanistic basis for the emergence of such behaviors. Finally, we should then seek to determine how such behaviors emerge in organismal brains. That is, are they hard-wired, or are they learned?

Here, we argue that self-awareness is essentially a learned behavior, that emerges in organisms with a sufficiently complex and highly integrated ability for associative memory and learning. The idea is that the organism's brain perceives the external world from a specific vantage point, and receives a continual set of inputs related to the organism as a result of this vantage point. This continual set of inputs related to the organism leads to the formation of neural pathways that define a set of associations related to various aspects of the organism, resulting in behavior consistent with self-recognition. We call this set of associations the organismal self-image.

This paper is organized as follows: In Section 2, we describe in further detail our speculation regarding the emergence of self-awareness. We consider various aspects of self-awareness, and speculate as to the possible mechanistic basis for their emergence. Specifically, we consider self-recognition, the emergence of the concept of "I/Me", and the emergence of an organism's awareness of being self-aware. This last phenomenon, in particular, has been the subject of intense philosophical and theological debate over the centuries. We also consider aspects of self-awareness that are apparently unique to humans, namely various existential and religious concepts such as the mind and soul, solipsism, the idea that reality is an illusion, and various conceptions of divinity. We also discuss the problem of tautologies that arises when considering the various existential questions associated with self-awareness, a problem that prevents an unambiguous resolution of the problem of self-awareness and the various associated existential issues. In Section 3, we discuss various experiments that could be used to manipulate self-awareness, as well as test some of the hypotheses presented here. Furthermore, we speculate on how certain kinds of metaphysical constructs, such as mind–body separation and the delocalization of mind in several bodies, may in principle be physically realizable. We conclude with a brief discussion of a number of additional issues related to self-awareness and computational neuroscience, such as the concept of qualia, and the role that set and logic theory could play in understanding various neural structures associated with higher cognitive functions in complex organisms.

2. Results

2.1. Self-awareness as a learned behavior

All sufficiently complex planetary life has a brain which processes the sensory input obtained by the organism's sensory organs. The five senses familiar to humans are sight, hearing, touch, taste, and smell (otherwise known as the exteroceptive senses, which also include balance). In addition to these basic senses, which provide the organismal brain with external information about the world, neural connections throughout the body provide the brain with information about various parts of the body in which it sits (Tannenbaum, 2001), as well as the spatial relationships between various parts of the body (these are the interoceptive and proprioceptive senses).

A seemingly obvious point to note is that the organism's brain sits inside the organism's head, and that the organism's various sensory organs and internal sensory connections are located in specific parts of the body, which are wired to the organism's brain in a specific way. Therefore, the organism's brain is obtaining information about the world from a specific vantage point. This vantage point is such that the flow of information from the environment always contains a constant subset of information related to the organism. Using humans as an example, the human brain constantly receives visual information from the human in which it sits. This visual information consists of various parts of the human organism's body, such as hands, arms, tip of the nose, feet, legs, torso, etc. It should be emphasized that, from the perspective of the brain, the environmental inputs also include the "internal" sensory inputs, since these provide information about certain aspects of the material world, in this case coming from the organism itself.

Since this subset of "self-information" is constant, the organism's brain can re-wire itself in response to this information flow, until neural pathways are formed that respond to stimuli connected to the organism. The end result of this "self-learning" process is the formation of neural pathways that respond to inputs related to the organism, and therefore recognize the organism. The neural pathways are said to, then, store a kind of "self-image" that is associated with the organism. Inputs that are related to this self-image then trigger the relevant neural pathways, resulting in self-recognition. Given that it is believed that neural pathway formation is driven by a reward–punishment-based selection process, we argue that the emergence of an organismal "self-image" should be a self-reinforcing process that automatically extracts the subset of information flow associated with the organism itself.

It makes sense that self-awareness should be processed in regions of the brain that integrate many different sources of inputs together. First of all, such a region of the brain will have a stronger total signal strength connected to the organism. Furthermore, since the different sources of external inputs carry different aspects of information about the organism, these respective aspects will all be strongly correlated with one another. In an organism with a sufficiently powerful associative memory and learning ability, these various self-inputs will result in the formation of a highly correlated, multi-dimensional "super-input." Because sources of inputs that derive from the environment, and not the organism, are constantly changing, the strength of the multi-dimensional signal obtained from the organism becomes proportionally stronger as the number of inputs that are integrated together increases (in a rough analogy, the environmental inputs are like noise terms that are averaged away as more and more sources of inputs are integrated together).